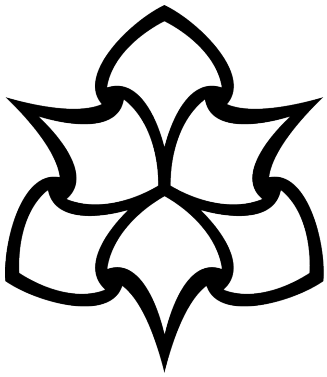

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2018, accepted March 9, 2018, date of publication March 20, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2817560

A Systematic Review of the Availability and Efficacy of Countermeasures to Internal Threats in Healthcare Critical Infrastructure

STEVEN WALKER-ROBERTS[1], MOHAMMAD HAMMOUDEH[1], AND ALI DEHGHANTANHA[2], (Senior Member, IEEE)

[1]School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester M1 5GD, U.K.
[2]Department of Computer Science, University of Sheffield, Sheffield S10 2TN, U.K.

Corresponding author: Mohammad Hammoudeh (m.hammoudeh@mmu.ac.uk)

ABSTRACT Insider attacks are becoming increasingly detrimental and frequent, affecting critical infrastructure at a massive scale. Recent attacks such as the U.K. National Health Service WannaCry ransomware attack which partly depends on internal users for initial infection highlight the increasing role of the malicious insiders in cyber-attack campaigns. The objective of this research is to ascertain the existing technological capability to mitigate insider threats within computer security systems by way of a mixed-method systematic review. Evidence was acquired from major sources of mainstream and grey literature by analyzing about 300 000 papers. Crude aggregated results were analyzed across the literature, and the results were TPR 0.75, FPR 0.32, σ 0.24 and 0.36, respectively, and σ^2 0.06 and 0.13, respectively. In totality, the literature evidence suggests that there is high heterogeneity across crude data indicating that the effectiveness of security measures varies significantly. No solution is able to totally mitigate an insider threat. Themes when set against that data suggest that most, if not all, security measures require breaches to occur before an analysis of malicious activity can prevent it in future through recall. Such a reactive approach is not effective to protect our critical infrastructure including our healthcare systems. Consequently, there is a major theoretical shortfall in current cyber defence architecture.

INDEX TERMS Critical infrastructure security, personal data safety, healthcare, data breach, insider threat, meta-data, sabotage, systematic review, thematic analysis, unprivileged, untrusted, zero trust.

I. INTRODUCTION

There are more data breaches reported now every year than one could care to count [1]. A significant percentage of these breaches were experienced in critical national infrastructure, which includes public health sector, power, communications, transportation, oil and gas, and financial institutions. In wartime, these are often designated as preferred military targets, which when compromised, will cause public panic, disconnection of communications and disruption to transportation. Today, the world is in the infancy phase of electronic warfare. Cyber attacks offer the ability to destroy or disrupt infrastructure targets remotely and anonymously, in very stealthy ways. In the healthcare sector for example, many corporations are interlinked with the government, and hence, data breaches can have a destructive impact on an entire nation. Citizens confidence and the economy will be affected by data compromise as many companies that operate public services also have other government contracts and interactions.

One of the major contributing factors to the increasing prevalence of ''data disasters'' is the inability to resolve the age-old problem of what happens when either (i) a person trusted to use a computer system betrays its owner to commit cybercrimes, (ii) a hacker casually makes his way past a firewall and sits behind it for some time committing cybercrimes to the almost certain ignorance of the system administrators. It is suggested by some commentators such as [2] that security often works like an onion with layers upon layers of security zones. It suggests that all it takes is for an inside threat to slip between security zones and they will become virtually undetectable, particularly if novel threats.

Catastrophic data breaches are becoming the story of the day increasingly often. Most recently, as at the date of publication, was the Equifax data breach in which potentially

information on 143 million US citizens and 44 million British citizens was stolen by hackers in May-July 2017. Before that was the NHS cyberattack. There is the Ashley Madison Breach, the TalkTalk breach, the OPM breach, the CIA/NSA ''hacking tool'' leaks, the Yahoo data breach, the Sony data breach, the MySpace data breach and so on – these organisations are not small players. These particular breaches have been apparently focussed on theft, but it would be right to question what would happen if instead they chose to sabotage or intentionally compromise a system or infrastructure in such a way as to seriously endanger life. This is increasingly relevant due to the role of cyberwarfare in statecraft.

Where it concerns healthcare specifically, cybersecurity has the potential to threaten life very easily. Most NHS trusts in England and Wales have application services presented as web applications with various backing stores, the most common being static file stores and databases. These are served at desktops, mobile devices and on ubiquitous devices (including medical equipment such as patient monitors). If these services become compromised or are successfully attacked, then critical internal infrastructure services such as access to laboratory results, radiography and real-time patient physiological information will be unavailable. Medical devices themselves can also be compromised, for example by DDoS on the wifi networks which they use to communicate with central monitoring stations. There is the additional danger of data theft owing to the exchange of data across so many devices. The Verizon VCDB dataset shows that over 1200 serious attacks were directed specifically towards healthcare infrastructure, which it identifies to be an increasing trend [3]. The recent NHS WannaCry attack is said by the National Audit Office to be the largest of its kind affecting a healthcare organisation in recorded history [4]. Embarrassingly, that report confirms that the NHS was not even a specific target, but had failed to comply with policy directions for the improvement of infrastructure and was still widely using Windows XP that was at that time no longer supported by Microsoft.

In a critical infrastructure context, the problem was that policy was clearly disengaged from front-line practice in the NHS [4]. One industrial research report illustrates that on more than 51% of occasions, the blame for a cybersecurity breach is negligent internal [5]. Similarly, another report projected that cybersecurity breaches were likely to cost healthcare providers potentially ''$300 billion'' in the future [6]. The most reported cause of cybersecurity breaches is negligence, therefore, it has to be questioned how this can be mitigated in practice. The National Audit Office identified, that had the WannaCry ransomware not been disrupted by coincidence when a cybersecurity analyst discovered a ''phone home'' mechanism by accident, then it is likely that significantly more devastation would have occurred [4]. Therefore, this scoping exercise must be conducted to understand the threat of internally-directed attacks in critical infrastructure such as in a healthcare setting like the NHS.

In the present threat climate, it is reasonable to question whether security breaches must be as a result of something more than *a failure to follow best practices and why existing measures are ineffective*. This mixed-method systematic review aims to investigate precisely that issue. A systematic review is an evidence-based literature review which goes beyond an ordinary review in rigorously assessing the quality of the literature using methods approved by the body of academic opinion. This approach was used because of the number of dogmatic practices in cybersecurity and little encompassing research which challenges that position as being unsatisfactory, it aims to be a fresh alternative to the typical survey of computer science literature which provides thorough critical analysis.

We have chosen the systematic review style to address the shortage of knowledge about effects of insider threats against security of critical infrastructure, particularly in the healthcare sector, because it is a highly approved academic scoping method within public health in the United Kingdom and abroad. This is owed mostly to the fact that systematic reviews are impartial and concise with adherence to a specific protocol. It amounts to an excellent tool for ''proving'' the state of the art as opposed to a subjective (potentially biased) literature summary in a survey context. The closest work to ours is the systematic literature review of insider threats offered by [7]. However, they have utilised a challenge metric which compounds potential differences affecting performance and effectiveness metrics for specified algorithms in Intrusion Detection and Prevention Systems (IDPS), thus harmonising data to fit a meta-analysis which would otherwise be inappropriate (this also introduces a substantial risk of bias). As a result, there is an opportunity to conduct an updated systematic review more relevant to internal threats. The objective of this systematic review is to ascertain the state of the art in computer security where the ability to mitigate insider threats within computer security systems is concerned in particular, especially as it relates to critical public infrastructure such as in the healthcare setting. It aims to extract data from the literature using mixed qualitative and quantitative methods. The quantitative data extracted will be explored in the context of qualitative themes in narrative synthesis lending itself to the mixed method extraction of data from studies. To achieve this aim, the following research questions will be answered:

1) To what extent are current technologies able to mitigate insider threats which abuse privilege?
2) What is the current research trend for insider threat mitigation?
3) What are the most effective methods of mitigating insider threats?

This systematic review is only concerned with the effectiveness of existing security measures to mitigate inside threats, the present research trend and the extent current technology is able to mitigate internal threats within a computer security system.

The rest of this paper is organised as follows: Section II presents the systematic review methodology. Section III presents the systematic review results. Section IV discusses the review results. Section V gives recommendation for practice and associated theoretical implications. Section VI concludes the paper and gives future research directions.

II. METHODOLOGY

The following databases were searched: ACM Digital Library, BASE (GreyLiterature), Collection of Computer Science Bibliographies, DANS (Grey Literature), dblp (Grey Literature), IEEE Xplore, JStor, OpenGrey (Grey Literature), ScienceDirect, Springer, Wiley, Zetoc (Grey Literature). There were two reviewers. The search returned 2577 results, of which 474 were duplicates, leaving the actual number of results at 2103. The search terms used were as follows: ((computer AND (misuse OR abuse)) OR (inside* NEAR threat), ((computer AND (misuse OR abuse)) OR (inside* NEAR threat)) AND (unprivileged OR trust OR privilege), (unprivileged). The literature search was intentionally cast wide to consider as many results as possible in connection with the research questions posed in this systematic review.

Grey literature was searched to avoid publication bias. All results were blinded as to publication status during the sifting phase. The review protocol is summarised in Figure 1 and a statistical summary of the results returned for each database search are included at Table 1.

TABLE 1. Results summary.

Database	# Results
ACM Digital Library	401
BASE (Grey Literature)	27
Collection of Computer Science Bibliographies	388
DANS (Grey Literature)	0
dblp (Grey Literature)	4
IEEE Xplore	95
JStor	0
OpenGrey (Grey Literature)	27
ScienceDirect	70
Springer	1531
Wiley	15
Zetoc (Grey Literature)	19
Total (474 duplicates)	2577

A. SELECTION OF STUDIES

The sifting phase is where each individual piece of literature was assessed against the inclusion and exclusion criteria and a decision made as to whether it should be excluded or not. The sifting process was divided into following six phases.

In the first phase, search results were filtered according to the inclusion and exclusion criteria, specifically the date and the academic field concerned. In this case, studies were selected from the past 9 years in the field of computing. The reason studies were not selected prior to 9 years ago is because of the rapidly developing state of the literature in that time which casts the relevance of earlier studies into doubt. More than 300,000 results were excluded at this stage.

In the second phase, all results were sifted based on apparent relevance to the research questions by title and abstract alone. 1195 results were excluded.

In the third phase, the remaining results were sifted on specificity by way of full reading. The full text article had to relate closely enough to the research questions posed in this systematic review. 720 results were excluded at that point.

In the fourth phase, all results were checked in detail for the presence of sufficient data which was appropriate in context to the research question posed (''effectiveness''). Many results were excluded because they measured only computing performance of a purported novel algorithm, not effectiveness. Thus, 96 results were excluded at this point.

In the fifth stage, all results were checked for quality using standardised testing tools for quantitative and qualitative research (see III-A). The results were further scored against set quality criteria within the protocol of the systematic review. A total of 22 results were excluded at this stage as having not met the minimum criteria of quality.

A total of 70 studies remained to be considered for inclusion by way of full critical analysis. Of the available 70 studies, 18 were included and the rest (52) were discarded either because they scored less than R4 for relevancy or had higher than B1 for risk of bias. The remaining 18 were fully analysed. None of the 18 remaining results were from grey literature sources. Though it was within the criteria that studies should have a high impact, some borderline studies were included to avoid bias despite being low impact. A full result set can be found in [8].

The assessment process was stringent to ensure that only the highest quality studies with solid findings were considered for mixed-method synthesis due to the risk that detail could otherwise be abstracted by the methodology or a poor complementary synthesis. Borderline studies were excluded which had relevance scores of R3 or bias scores of B2. This was to avoid any potential risk of bias within the data extraction and synthesis. Borderline cases which had low impact were still considered as a mitigation against the risk of publication bias.

B. QUALITY ASSESSMENT

Quality for the research studies was broadly assessed in four ways. Qualitative studies were critically appraised using the CASP tool [9] for qualitative research, which is a well-established method of qualitative critical appraisal. Quantitative studies were assessed using the SURE [10] critical appraisal tool which is a generic quantitative research assessment tool well-suited to the field of computer science due to heterogeneity of methodologies within studies. Any grey literature was assessed using the AACODS critical appraisal tool [11]. Following critical appraisal, studies were then judged against the quality criteria of the review itself as set down in the protocol (see Appendix).

C. DATA EXTRACTION

This is a mixed method systematic review undertaking Joanna Briggs Institute (JBI) approach [12]. This approach provides an evidence-based methodology for combining the results of qualitative and quantitative research. In this systematic review, included studies were analysed in both a quantitative and qualitative manner and so this is favourable. There were two reasons for this: (i) there were mixed qualitative and quantitative studies, though the majority were quantitative (ii) more data existed than the quantitative data provided in the studies alone.

The approach taken to analysing the quantitative and qualitative data contained in the studies was to take a two stage process. In the first stage, results and discussion were analysed for key findings and moments of importance to the objectives of those studies. These were converted to textual descriptions which were then further thematically analysed blinded as to the author or article title. The textual descriptions were coded and then themes were extracted from the codes emerging from the textual descriptions.

Then, from extracted quantitative data, common measures of results were identified and the figures extracted for further statistical analysis. Meta-analysis was planned but did not proceed because study methodology and sample sizes, along with factors affecting results, were too heterogeneous to safely perform a meta-analysis. Instead, the result measures were aggregated using a crude grand mean for True Positive Rate (TPR) and False Positive Rate (FPR) for those studies that provided those values.

III. RESULTS

A. EXTRACTED QUALITATIVE DATA

As described in Section II, all research papers were explored for moments of importance using the phenomenological approach of thematic analysis. Though the majority of the studies were not themselves qualitative, the JBI approach of textually describing key moments in the quantitative studies provided significant qualitative data which could then be used for thematic analysis. The textual descriptions were coded and then key themes extracted from recurring and similar codes. The textual descriptions and full thematic analysis can be viewed in [8]

A frequency chart of codes is provided in Fig. 2 and a statistical breakdown of the emergent themes in the thematic analysis in Table 2. It is apparent that the most common codes are: anomaly detection (8), comparison of user behaviour (8), machine learning (6), behaviour profiling (11), context dependent (5), low accuracy (4), malicious insider undeterred (4), algorithm optimisation (4) and improvement of algorithm (5).

The most common themes are: anomaly detection (11.5%), context dependence (14.9%), profiling (21.8%), accuracy (10.3%), scalability (13.8%), improvement of algorithm (10.3%).

The accuracy of the analysis is confirmed by the known fact that IDPS are the predominant mainstream utility for mitigating insider threats. Only first order themes were derived as these appeared sufficient in quantity and quality to address the research problems. Had the first order themes been subjected to second order thematic analysis, the resulting themes would have been too inclusive.

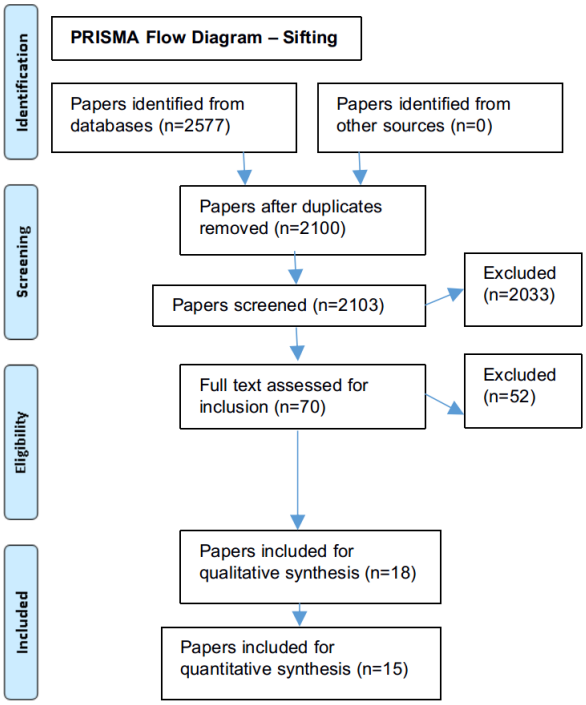

FIGURE 1. PRISMA flow diagram.

TABLE 2. Breakdown of theme occurrence.

Theme	%
anomaly detection	11.5
context dependence	14.9
profiling	21.8
accuracy	10.3
scalability	13.8
machine learning	6.9
undeterred	4.6
improvement of algorithm	10.3
controlling risk	5.7
TOTAL	100

B. EXTRACTED QUANTITATIVE DATA

All studies had result data, sample data and methodology extracted and placed onto a spreadsheet as in [8]. That data was then analysed for common measures. It is apparent from the spreadsheet used in the systematic review that 12 of 18 articles had a common measure of True Positive Rate (TPR), whilst 7 of 18 articles had a common measure of False-Positive Rate (FPR).

It was noted from the results that a common way of assessing TPR and FPR together is by constructing a Receiver Operating Characteristic (ROC) curve in which a given ROC curve

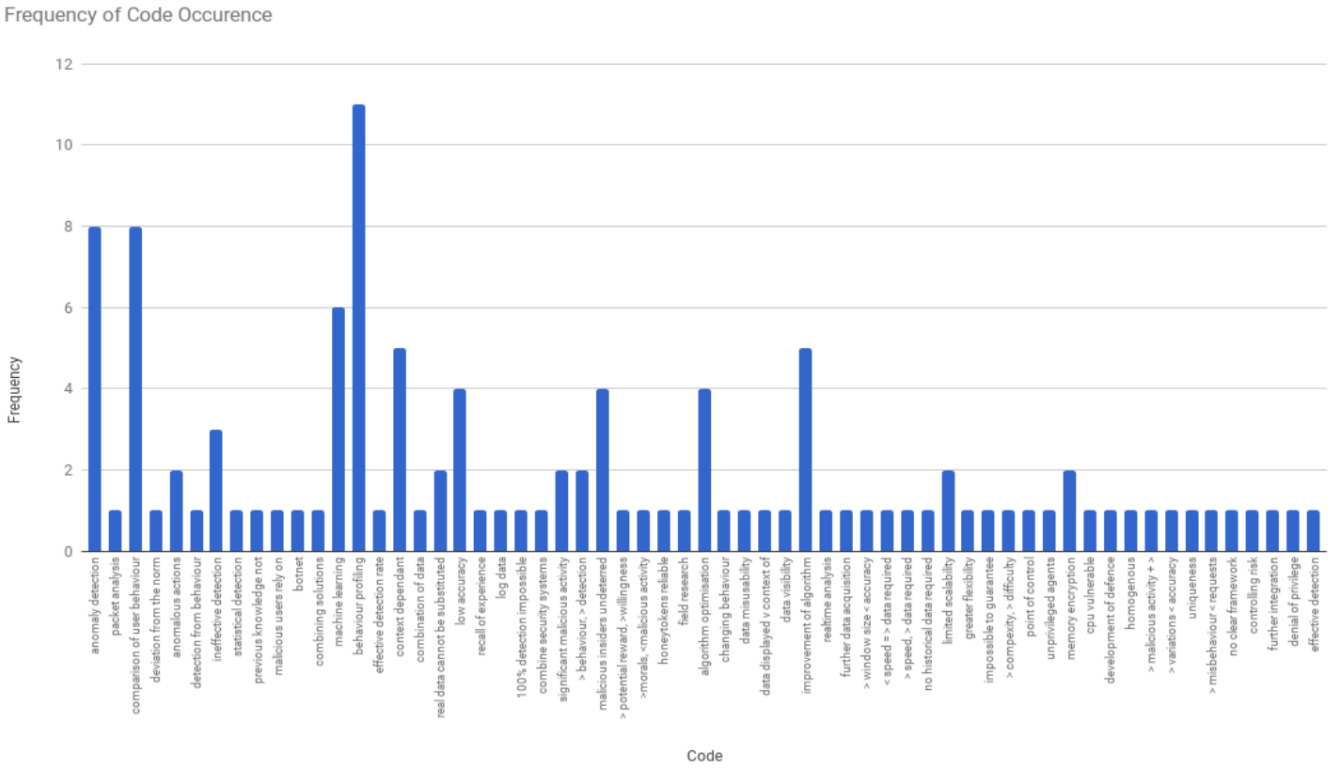

FIGURE 2. Frequency of code occurrence.

generated from reported results in individual studies, or their crudely aggregated results, could be used to assess study results against the theoretical ideal of TPR 1.0 and FPR 0.0. In this case, it was justifiable to aggregate study results because not all studies report TPR and FPR together. Since study method and sample size are heterogeneous, it is not possible to weight means together or analyse risk ratios or odds ratios for known influential factors in each study. Thus meta-analysis is not feasible nor is a weighted average TPR and FPR in respect of each study since this would completely lose the resolution of the data.

The crudely aggregated grand mean was taken from mean values of the lowest TPR/FPR and the highest TPR/FPR reported in each study. Some studies did not provide detailed data but instead an author-calculated mean TPR/FPR. Thus, it was reasonable to aggregate all mean TPR/FPR values. Mean lowest and highest reported TPR/FPR values were used to construct an aggregated ROC curve in respect of 13/18 studies included in the systematic review. It should be borne in mind that quantitative non-inclusivity is 27.78% in respect of resultant values. This is, however, mitigated by qualitative analysis.

The aggregated grand mean was TPR 0.75 and FPR 0.32. Euclidean distance from the ideal is 0.25 for aggregated mean TPR and 0.32 for aggregated mean FPR, which is numerically significant. Mean lowest reported TPR and FPR were 0.57 and 0.17 respectively. Highest reported TPR and FPR were 0.84 and 0.36 respectively. The mean range for

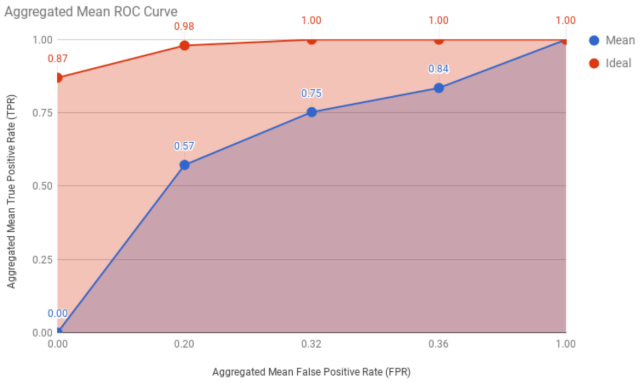

FIGURE 3. ROC curve for mean and ideal values of FPR and TPR.

reported TPR and FPR values were 0.31 and 0.17 respectively. Variances among mean TPR and FPR were 0.06 and 0.13 respectively. Standard deviations among mean reported TPR and FPR were 0.24 and 0.36 respectively.

The detailed TPR and FPR values for each study are listed in Table 3. The ROC curve for ideal and actual grand mean values are presented in Fig. 3. It is notable that actual results are markedly below the theoretical ideal in the ROC graph.

IV. DISCUSSION

The significance of an inside threat was much dependent on the context in which the user acts maliciously. Both studies conclude, on the basis of apparently sound findings, that

TABLE 3. TPR and FPR values for included studies.

Reference	Non-Ideal TPR	Low TPR	Mean TPR	High TPR	Range TPR	Ideal TPR	Ideal FPR	Range FPR	Low FPR	Mean FPR	High FPR	Non-Ideal FPR
Gafny et. al (2010)	0	–	0.92	–	–	1	0	–	–	0.03	–	1
Shabtai et. al. (2016)	0	0.71	0.86	1	0.29	1	0	–	–	–	–	1
Baracaldo and Joshi (2012)	0	0.2	0.45	0.7	0.5	1	0	–	–	–	–	1
"Hussain, Sallam & Bertino (2015)"	0	–	–	–	–	1	0	0.3	0.38	0.53	0.68	1
Yu (2011)	0	–	0.99	–	–	1	0	–	–	–	–	1
Bose et. al (2017)	0	–	0.5	–	–	1	0	–	–	0.92	–	1
Nasa and Varjana (2014)	0	0.2	0.3	0.4	0.2	1	0	–	–	0.05	–	1
Legg et. al (2017)	0	–	1	–	–	1	0	–	–	0.58	–	1
Alotibi et. al. (2016)	0	0.66	0.82	0.98	0.32	1	0	–	—	—	—	1
Mohan et. al. (2015)	0	0.89	0.95	1	0.29	1	0	0.03	0.01	0.03	0.04	1
Chen et. al. (2012)	0	0.5	0.55	0.6	0.29	1	0	–	–	0.1	–	1
Sun et. al. (2016)	0	0.92	0.96	1	0.09	1	0	–	–	–	–	1
Liu et. al. (2011)	0	0.5	0.75	1	0.5	1	0	–	–	–	–	1
Mean		0.57	0.75	0.84	0.31			0.17	0.2	0.32	0.36	
Variance		0.08	0.06	0.06	0.02			0.04	0.07	0.13	0.2	
SD		0.28	0.24	0.24	0.14			0.19	0.26	0.36	0.45	
Ideal		1	1	1	1			0	0	0	0	

context was extremely subjective and thus it was impossible to adjust detection systems to be more or less sensitive to a particular context-based indicator. Both studies identify that the future research direction should be focussed towards discriminating context, perhaps by combining multiple IDPS technologies together to narrow the subjectivity of malicious contexts.

Both of these studies had a moderately high FPR and so the quantitative data reported in each study tends to support, rather than contradict, the qualitative data extracted from those studies. [13], which proposes the high impact SNAD algorithm, identifies that the algorithm struggles to identify malicious activity where expected user behaviour is extremely homogeneous. In support of the above studies, it finds that a future research direction will inevitably be to develop semantic models of user behaviour in order to underline context in malicious activity in insiders.

A. PROFILING

Profiling also features strongly within the literature at 21.8% occurrence within thematic analysis. Of 18 studies, 11 highlighted profiling as an important element. This confirms that roughly 2/3 of the studies included in the systematic review use historic behavioural data to examine anomalies potentially disclosing an inside threat.

The TPR values among methods using historic profiling are higher as are FPR values, when compared to non-profiling methods of IDPS. Non-profiling methods of IDPS appear to identify a numerically significant lower reported value of TPR and FPR. It is difficult to explain with precision the reason for this, but it is likely on the evidence available within the systematic review that this is accounted for by the fact that an IDPS which holds no historic data can not be prejudiced by historic data so as to exclude it at some future time.

A possible attack surface of profile-based IDPS technologies is that a malicious insider is either able to skew the historic profile to repudiate their activity or they are able to normalise malicious activity. Whilst non-profiling IDPS technologies report a lower TPR and FPR, they are not vulnerable to this phenomenon. A number of studies within the systematic review such as [14] identify that improved signature generation is an area of future works for this reason.

B. ACCURACY

Accuracy features moderately as a theme within the literature with an occurrence at 10.3%.

It is a major problem within IDPS systems. In thematic analysis, 9 codes were related to serious accuracy problems within IDPS systems. The most inaccurate were the alarm-based anomaly detection system investigated by [15], with less than a 20% detection rate, and the RADISH system in [16] with a 50% detection rate. The rates of detection do not represent a poor study outcome or indeed a poor study (this would have been a publication bias), however it does present the need for significant further investigation.

Of the 12 studies the reported TPR and FPR values, it is apparent that their aggregated values fall significantly below the ideal ROC curve in Fig. 3. Given the context of the systematic review in investigating internal threats, this feature is important because as had already been described, one malicious activity is enough to be catastrophic.

The inability of any study to reliably prove a 100% TPR suggests that IDPS is not generally designed to prevent the types of inside malicious activity that are resulting in major data breaches. The tabulated range of the TPR and FPR values reported in studies appears to confirm the same problem.

C. SCALABILITY

The theme of scalability was moderately emergent within included studies at 13.8%. Studies included in the systematic review reported a mixture of scalability issues. These included the need for greater flexibility in scaling up detection resources, issues with the flexibility of revocation of access and the ability of IDPS systems to cope with much larger volumes of data for analysis.

References [13] and [17] particularly highlight that when IDPS systems are scaled, naturally their TPR and FPR rates are adjusted, often because of increased inaccuracy at higher volumes. References [16] and [18] suggest that the only real way of addressing scalability issues is by combining multiple security methods to mitigate the effect of scalability. However, those studies do stop short of testing this approach and identify this as an area of future work, thus it is not possible to conclude with any level of precision whether taking that modified approach would be effective.

In particular, the scalability issues identified in the literature create concerns where big data and cloud services are concerned. If an IDPS can not be scaled up, then it is reasonable to question whether it can mitigate threats in a complex distributed computer system where system activity may exponentially increase over time. Despite this, there does not appear to be any data within these studies to prove that there would definitely be a scalability problem in respect of each approach taken.

D. MACHINE LEARNING

Machine learning features less within the literature at a 6.9% theme occurrence. This is still significant. From the studies included in the systematic review, it is reasonable to conclude that machine learning in IDPS is an emerging academic interest. Studies take a mixed approach to application of machine learning in IDPS systems.

Reference [19] appears to be the earliest article of all studies included, which uses petri nets to classify whether user activities are taking place in an acceptable order. Conversely, [20] uses finite state machines to create a fuzzy model of malicious activity which can then be subject to binary decisions based on set threshold. References [13] and [21] are studies which also apply inductive machine learning models to determine the definition of anomalous behaviour. Other studies use k-nn and k-decision tree machine learning algorithms, such as in the RADISH system [16].

It is highly notable that in every study except SNAD [13], the detection rate is extremely high, with high values of TPR and low values of FPR. The TPR and FPR values are quite close to ideal, with a low Euclidean distance in respect of the same. Though this is observed, of 5 studies using machine learning, only two use real-world data to test the machine learning algorithms posed. It is therefore not possible to conclude with any degree of precision how well machine-learning based IDPS would tolerate real world malicious insider activity.

E. MALICIOUS INSIDERS UNDETERRED

The theme of malicious users being undeterred represents a small but statistically significant occurrence at 4.6%.

In quantitative studies, particularly [22], it was proven statistically that even though users said they would act differently in the knowledge of honeytokens, those that did know about the honeytokens were not at all deterred. There are serious drawbacks with the approach taken in this study because whilst the study was controlled and participants blinded, as the study points out employees who could face disciplinary action and potentially prosecution would treat the situation differently to students who know the exercise is a simulation.

In the MITRE trade secrets study, [23], employees reported that they would react differently if they were aware that their malicious behaviour was intercepted. Regrettably, this study does not test these results further and so it is not possible to fully compare this study with [22]. It is noteworthy that only 4/173 malicious actors were deterred in [22], in the MITRE study users took significant evasive action to hide their malicious activities. The issue likely to identify with these studies is that a malicious insider in the real world may behave very differently and so only limited weight can be given to the information conveyed in these studies.

F. IMPROVEMENT OF ALGORITHM

Improvement of algorithm features as an important theme among included studies with an occurrence of 10.3%. Of 18 studies included in the systematic review, 5 ([21], [22], [14], [24], [25]) identified discrimination of malicious inside threats and the need for less intervention by an administrator as areas for significant improvement. Since the study data does not explicitly relate to these conclusions, it is not possible to conclude with certainty whether the authors in these studies took an accurate position. However, it appears on the basis of aggregated TPR and FPR values that these conclusions may be true of all included studies.

If all included studies require improvement in the same manner, this could explain the difference between the ROC curves in Fig. 3. It could provide substantiation to the idea that IDPS is not designed to deal with novel inside threats.

G. CONTROLLING RISK

The theme of controlling risk features as a small but statistically significant occurrence at 5.7%.

References [21] and [26] both suggest and propose that risk can be controlled using a risk-reward approach. When a user does a malicious act their trust rating is downgraded until access is entirely denied. When a user engages in normal use, their trust rating is restored.

This approach is useful, but because it is longitudinal it may take time to detect malicious activities. It only takes a single malicious activity to be catastrophic. In addition, a malicious user may abuse the disposition of risk analysis by normalising their behaviour as they engage in malicious activities to repudiate their malicious acts. This is a serious drawback with risk analysis alone. Another manner of controlling trust is described in [27], but in the context of such a dangerous exploit could be the only real solution.

It is important that of all 18 studies included, there is little consideration of controlling risk which one could rightfully conclude would be important since the majority of studies confirm that a serious drawback in every case is an inability to detect 100% of malicious activity and mitigate it. Clearly risk control is an area where significant future research is required.

V. COUNTERMEASURES TO COMBAT GROWING NUMBERS OF CRITICAL INFRASTRUCTURE SECURITY BREACHES

It is submitted that the source of the ''straw that broke the camel's back'' security breaches is not within necessarily unapproved security practices or software failures. This systematic review highlights that the issue is much more serious.

The existing security technologies most commonly deployed today require statistical induction and are often heuristic in the absence of ''experiential knowledge'' of a potential threat. Thus, it is suggested that the only real mitigation is a complete redesign of computer infrastructures to not only make all resources immutable, but to remove the ability of an attacker to navigate resources across different infrastructure layers. This, it is submitted, is the only way of preventing compromise in a threat climate where a single system event can lead to catastrophic outcomes.

It is recommended that because the majority of security measures can not by themselves mitigate a catastrophic inside threat to a security system, multiple security measures must be used together to moderate otherwise substantial risk of catastrophe.

The US Department of Health and Human Resources that identifies in the OCR Breach Report that the majority of incidents were data disclosure incidents, mostly operator error, or were internal attacks that remained undetected for a significant period of time resulting in loss and damage [28].

It appears from the literature that the best approach to be taken is to incapacitate a malicious insider by removing data visibility and locking out permissions entirely so that internal privilege can not be abused. In an increased threat climate which the literature suggests cannot be entirely mitigated, it is extremely important that system administrators do not rely on the automation some technologies provide and remain alert to unusual activity that may not be automatically alerted to them. System and software design should take into account the need to mitigate the risk of an internal threat starting at the very lowest level.

The literature body generally as included within this systematic review takes a common focus towards IDPS with few studies focussing on other ways of mitigating insider threats. This represents a deficiency in the scope of active research in the fields of computer science and computer security.

Taken together, all studies confirm that 100% detection of malicious insider threats are not possible and that to some extent, malicious insiders are not deterred by in-place security measures. This is a very important feature within the literature.

In recent years, it is apparent that there is an increasing trend within the literature towards predictive behavioural modelling to identify early malicious activity before a catastrophic data breach, though it will be clear that these behavioural systems take time to work and are therefore inappropriate in dealing with zero day or other novel inside threats.

It is important to note that the literature consensus appears to be that ''malicious insider'' and ''inside threat'' are very poorly defined and are applied loosely to mean a person, whereas in practice an insider could be an outsider with privileged internal access to a computer system.

Machine learning in recent years has become highly prominent within the literature, accounting for a large number of included studies. These studies have the highest levels of accuracy in terms of TPR and FPR. However, the majority of machine learning studies failed to test real-world datasets.

Accuracy features prominently in all included studies, with those testing real-world data appearing to perform the most poorly. The reason for this is not particularly clear but may be because of optimistic modelling. On the whole it is clear that no reported technique within the included literature base can mitigate 100% of inside threats, and thus cannot prevent a prospective single fatal breach.

The majority of studies identify that proposed algorithms need significant algorithm optimisation by way of improved signature generation, improved moderation of risk and improved scalability. In totality, the literature suggests that there is no way to mitigate ''knockout'' data breaches which could effectively destroy an organisation, cause serious data loss or pose a significant threat to personal safety as a result of a sabotage of critical infrastructure. This is notable.

Theoretically, it appears that the only way to entirely mitigate an inside threat is to entirely remove privilege. It is possible that by controlling the degrees of freedom associated with specific permissions and data visibility, a malicious insider can be ''sandboxed''.

It may be possible to develop a model based upon the degrees of freedom of a computing resource and a potential malicious insider. The present literature base as included within this systematic review suggests that the approach that needs to be taken is to treat all users as a threat purely because it may not be possible to identify a malicious insider until it is too late.

VI. CONCLUSION AND FUTURE WORK

This systematic review was limited in that meta-analysis was not possible due to heterogeneity across studies. This is a very important remark because the resulting issue was that quantitative data could only be synthesised within 72.22% inclusivity. There is therefore substantial risk that quantitative data may have been abstracted.

Whilst a qualitative methodology was applied for exactly that reason, qualitative data may not have made up for the absence of quantitative data which provides a temporal dimension to the data. Due to the need for study synthesis to be solidly founded on very reliable data and methodology in order for qualitative analysis to be useful, only 18 studies could be used due to issues surrounding quality and the lack of quantitative data which was reliable and relevant. This may have excluded a large number of potentially relevant studies in a potentially less stringent protocol. Too many studies focussed on performance not reliability.

The role of negligent insiders in critical healthcare infrastructure is only becoming more apparent. Thus, the need for improved technology needs to be balanced against the need for user education and policy centred around the user that exposes critical healthcare infrastructure to catastrophe.

Significant work needs to be undertaken to create more effective IDPS techniques. Further work also needs to be undertaken to create a model of threat mitigation which takes

TABLE 4. Systematic review protocol.

Title of Review
The escalating role of inside threats in computer security breaches: a mixed-method systematic review of the availability and efficacy of inside threat mitigation approaches.
Objective
To ascertain the state of the art in computer security where the ability to mitigate inside threats within computer security systems is concerned in particular.
Research Questions
RQ1: To what extent are current technologies able to mitigate insider threats which abuse privilege? RQ2: What is the current research trend for insider threat mitigation? RQ3: What are the most effective methods of mitigating insider threats?
Hypothesis
It is hypothesised that most security mechanisms reported in the literature for mitigating inside security threats will reveal (i) a high prevalence of the threat (ii) an inability to deter the threat (iii) no mitigation methods with an acceptable standard of mitigation. This hypothesis is taken with the view that only one computer security breach is all that is necessary for a catastrophic resulting event to occur.
Reviewers
Primary reviewer and writer: Steven Walker-Roberts, Computer Scientist, Manchester Metropolitan University (LLB, MSc). Second reviewer: Mohammad Hammoudeh, Senior Lecturer in Computer Security, Manchester Metropolitan University (PhD) Third reviewer: Ali Dehghantanha, Lecturer in Cyber Security and Forensics, University of Salford (PhD)
Methodology
M1: Mixed systematic literature review - consider qualitative and quantitative studies This systematic review uses the JBI approach to conducting mixed method systematic review. The reason is because pilot searches revealed that quantitative data did not contain enough information to answer the research questions, whilst at the same time, qualitative research was sparse and many quantitative studies contained a significant amount of qualitative data, both experimental and non-experimental, which is important to consider. M2: Convert quantitative to qualitative research by creating thematic summary of qualitative studies The JBI approach provides a method of converting quantitative studies to qualitative data by extracting important, well-established facts from those studies and converting them into accurate non-biased textual descriptions. M3: Weight qualitative research using thematic analysis of methodology, data, results and synthesis In this review, those textual summaries are then mined and thematically analysed in order to provide a discrete qualitative synthesis using specific text bodies: methodology, data, results and synthesis. Only textual summaries which were factually established by the subject matter and methodology of the study are to be analysed. M4: Identify possible aggregations of data for further possible meta-analysis of original quantitative data extracted Extracted data is to be aggregated, where the measure of results is the same. If insufficient studies exist which pass the sifting stage and are aggregable then no further quantitative aggregation will be performed. If aggregable data exists, but little information exists as to study characteristics or methodology it too homogenous, then data will only be aggregated and not meta-analysed. M5: Synthesis based on emergent qualitative evidence and any aggregations of homogenous quantitative data Both the synthesised quantitative data and qualitative data will be discussed in narrative synthesis with reference to both datasets, the wider literature discourse and within the context of the themes identified by qualitative synthesis.
Search Methodology
Search Terms
((computer AND (misuse OR abuse)) OR (inside* NEAR threat) ((computer AND (misuse OR abuse)) OR (inside* NEAR threat)) AND (unprivileged OR trust OR privilege) (unprivileged)
Databases
ACM Digital Library, BASE (Grey Literature), Collection of Computer Science Bibliographies, DANS (Grey Literature), dblp (Grey Literature), IEEE Xplore, JStor, OpenGrey (Grey Literature), ScienceDirect, Springer, Wiley, Zetoc (Grey Literature)
Sifting Process
C1: Range Studies will initially be sifted on date range (studies from the last 9 years), field (computer science) and originality (must be original research studies).
C2: Relevance Title and abstract will be screened for relevance to the research questions posed.

TABLE 4. *(Continued.)* **Systematic review protocol.**

C3: Specificity Full reading to check that the research study relates closely enough to the research questions and insider threats to computer security specifically.		
C4: Data Studies searched for adequate data related to the research questions which is well-supported by methodology.		
C5: Study Quality Study assessed as a whole using CASP qualitative research quality assessment tool and the SURE quantitative research quality assessment tool. Studies then assessed further against additional quality criteria of this systematic review. Remaining studies considered for inclusion.		
C6: Inclusion Study assessed against inclusion criteria. Any study not meeting the full inclusion criteria is sifted.		
Study Criteria		
Inclusion Criteria: I01 Quantitative or qualitative original research study I02 Must address the efficacy of existing security measures I03 Experimental methods must address penetrability I04 Focussed on malicious actors on the inside of a computer system. I05 Must explore specific security vulnerabilities I06 Must be a technical paper		
Exclusion Criteria: E01 Non-peer reviewed or secondary research E02 Not related closely enough to insider computer security threats E03 Risk of bias E04 Not within the last 9 years E05 Lacks academic rigour or no data to analyse from a new study E06 Irrelevant or not in English		
Quality Criteria: Q01 Sufficient sample size Q02 Sufficient protection from bias Q03 Appropriate methodology Q04 Ethically sound Q05 Academically rigorous Q06 Sufficient research impact	Relevance: R1: No Relevance R2: Minimal Relevance R3: Acceptable Relevance R4: Highly Relevant	Bias B1: No Risk B2: Low Risk B3: High Risk
Record of Findings		
Findings will be recorded within a spreadsheet on Google Drive for computed analysis and so that, when the research is published, it can be viewed transparently to understand the systematic review further.		

into account the unknown malicious insider whom only need commit himself to one activity for it to be catastrophic. Additional work also needs to be undertaken to understand the nature of inside threats so that new technologies can be developed around potential further findings.

APPENDIX

See Table 4.

REFERENCES

[1] Verizon RISK. (Oct. 2017). *VCDB/Yearly.Png at Master · vz-Risk/VCDB*. Accessed: Nov. 2, 2017. [Online]. Available: https://github.com/vz-risk/VCDB/blob/master/figure/yearly.png

[2] S. Broderick, "Firewalls—Are they enough protection for current networks?" *Inf. Secur. Tech. Rep.*, vol. 10, no. 4, pp. 204–212, Jan. 2005. [Online]. Available: http://dx.doi.org/10.1016/j.istr.2005.10.002

[3] Verizon RISK. (Oct. 2017). *vz-Risk/VCDB: Veris Community Database*. Accessed: Oct. 26, 2017. [Online]. Available: https://github.com/vz-risk/VCDB

[4] National Audit Office. (Oct. 2017). *Investigation: Wannacry Cyber Attack and the NHS (Summary)*. Accessed: Feb. 7, 2018. [Online]. Available: https://www.nao.org.uk/wpcontent/uploads/2017/10/Investigation-WannaCry-cyber-attack-andthe-NHS-Summary.pdf

[5] Sans Institute. (2014). *New Threats Drive Improved Practices: State of Cybersecurity in Health Care Organizations*. Accessed: Feb. 7, 2018. [Online]. Available: https://www.sans.org/readingroom/whitepapers/analyst/threats-drive-improved-practices-statecybersecurity-health-care-organizations-35652

[6] Accenture. (2015). *The $300 Billion Attack—Accenture*. Accessed: Feb. 7, 2018. [Online]. Available: https://www.accenture.com/t20171221T005341Z

[7] I. A. Gheyas and A. E. Abdallah, "Detection and prediction of insider threats to cyber security: A systematic literature review and meta-analysis," *Big Data Anal.*, vol. 1, p. 6, Aug. 2016. [Online]. Available: https://doi.org/10.1186/Fs41044-016-0006-0

[8] S. Walker-Roberts. (Oct. 2017). *Systematic Literature Search*. Accessed: Nov. 3, 2017. [Online]. Available: https://drive.google.com/open?id=1b0YA6AJ5waHkCkp2UGAlP15 LTbyC8cori4Mw pKyA4

[9] K. Hannes, C. Lockwood, and A. Pearson, "A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research," *Qualitative Health Res.*, vol. 20, no. 12, pp. 1736–1743, 2010.

[10] Specialist Unit for Review Evidence. (Oct. 2017). *Critical Appraisal Checklists—Specialist Unit for Review Evidence—Cardiff University*. Accessed: Nov. 3, 2017. [Online]. Available: https://www.cardiff.ac.uk/specialist-unit-for-review-evidence/resources/critical-appraisal-checklists

[11] Flinders University. (Oct. 2017). *Aacods Checklist*. Accessed: Nov. 3, 2017. [Online]. Available: https://dspace.flinders.edu.au/jspui/bitstream/2328/3326/4/AACODS_Checklist.pdf

[12] Joanna Briggs Institute. (Oct. 2017). *JBI Reviewer's Manual*. Accessed: Nov. 3, 2017. [Online]. Available: https://reviewersmanual.joannabriggs.org/

[13] Y. Chen, S. Nyemba, W. Zhang, and B. Malin, "Specializing network analysis to detect anomalous insider actions," *Secur. Informat.*, vol. 1, p. 5, Dec. 2012. [Online]. Available: https://doi.org/10.1186/F2190-8532-1-5

[14] S. R. Hussain, A. M. Sallam, and E. Bertino, "DetAnom: Detecting anomalous database transactions by insiders," in *Proc. 5th ACM Conf. Data Appl. Secur. Privacy (CODASPY)*, 2015, pp. 25–35. [Online]. Available: https://doi.org/10.1145/F2699026.2699111

[15] P. M. Nasr and A. Y. Varjani, "Alarm based anomaly detection of insider attacks in SCADA system," in *Proc. Smart Grid Conf. (SGC)*, Dec. 2014, pp. 1–6. [Online]. Available: https://doi.org/10.1109/Fsgc.2014.7090881

[16] B. Böse, B. Avasarala, S. Tirthapura, Y.-Y. Chung, and D. Steiner, "Detecting insider threats using RADISH: A system for real-time anomaly detection in heterogeneous data streams," *IEEE Syst. J.*, vol. 11, no. 2, pp. 471–482, Jun. 2017. [Online]. Available: https://doi.org/10.1109/Fjsyst.2016.2558507

[17] A. Liu, J. Chen, and L. Yang, "Real-time detection of covert channels in highly virtualized environments," in *Critical Infrastructure Protection V*. Berlin, Germany: Springer, 2011, pp. 151–164.

[18] K. I. Santosa, C. Lim, and A. Erwin, "Analysis of educational institution DNS network traffic for insider threats," in *Proc. Int. Conf. Comput., Control, Informat. Appl. (IC3INA)*, Oct. 2016, pp. 147–152. [Online]. Available: https://doi.org/10.1109/Fic3ina.2016.7863040

[19] M. Chagarlamudi, B. Panda, and Y. Hu, "Insider threat in database systems: Preventing malicious users' activities in databases," in *Proc. 6th Int. Conf. Inf. Technol., New Generat.*, Apr. 2009, pp. 1616–1620. [Online]. Available: https://doi.org/10.1109/Fitng.2009.67

[20] Y. Yu. (2011). *Anomaly Intrusion Detection Based Upon an Artificial Immunity Model*. [Online]. Available: https://doi.org/10.1145/F2016039.2016075

[21] M. Gafny, A. Shabtai, L. Rokach, and Y. Elovici, "Detecting data misuse by applying context-based data linkage," in *Proc. ACM Workshop Insider Threats-Insider Threats*, 2010, pp. 3–12. [Online]. Available: https://doi.org/10.1145/F1866886.1866890

[22] A. Shabtai, M. Bercovitch, L. Rokach, Y. Gal, Y. Elovici, and E. Shmueli, "Behavioral study of users when interacting with active honeytokens," *ACM Trans. Inf. Syst. Secur.*, vol. 18, no. 3, Feb. 2016, Art. no. 9. [Online]. Available: https://doi.org/10.1145/F2854152

[23] D. Caputo, M. Maloof, and G. Stephens, "Detecting insider theft of trade secrets," *IEEE Security Privacy*, vol. 7, no. 6, pp. 14–21, Nov. 2009. [Online]. Available: https://doi.org/10.1109/Fmsp.2009.110

[24] G. Alotibi, N. Clarke, F. Li, and S. Furnell, "User profiling from network traffic via novel application-level interactions," in *Proc. 11th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Dec. 2016, pp. 279–285. [Online]. Available: https://doi.org/10.1109/Ficitst.2016.7856712

[25] R. Mohan, V. Vaidehi, A. K. A, M. M, and S. S. Chakkaravarthy, "Complex event processing based hybrid intrusion detection system," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–6. [Online]. Available: https://doi.org/10.1109/Ficscn.2015.7219827

[26] N. Baracaldo and J. Joshi, "A trust-and-risk aware RBAC framework: Tackling insider threat," in *Proc. 17th ACM Symp. Access Control Models Technol. (SACMAT)*, 2012, pp. 167–176. [Online]. Available: https://doi.org/10.1145/F2295136.2295168

[27] B. Hopkins, J. Shield, and C. North, "Redirecting DRAM memory pages: Examining the threat of system memory hardware trojans," in *Proc. IEEE Int. Symp. Hardw. Oriented Secur. Trust (HOST)*, May 2016, pp. 197–202. [Online]. Available: https://doi.org/10.1109/Fhst.2016.7495582

[28] U.S. Department of Health and Human Services. (2018). *Ocr Breach Report*. Accessed: Feb. 7, 2018. [Online]. Available: https://ocrportal.hhs.gov/ocr/breach/breach report.jsf

STEVEN WALKER-ROBERTS is currently an Honorary Research Fellow and a Sessional Lecturer with Manchester Metropolitan University. He is also an Academic Lawyer. He is also a Public Policy Adviser to the U.K. Government, having made significant contributions to home affairs and the criminal law. His research interests include national security, statecraft, and zero trust security. Other interests include technology law and the practical implementations of zero trust security systems to mitigate complex security threats.

MOHAMMAD HAMMOUDEH is currently the Head of the MMU IoT Laboratory and a Senior Lecturer in computer networks and security with the School of Computing, Math and Digital Technology, Manchester Metropolitan University. He has been a researcher and publisher in the field of big sensory data mining and visualization. He is a highly proficient, experienced, and professionally certified cybersecurity professional, specializing in threat analysis, and information and network security management. His research interests include highly decentralized algorithms, communication, and cross-layered solutions to Internet of Things, and wireless sensor networks.

ALI DEHGHANTANHA (GS'07–M'12–SM'16) is currently a Senior Lecturer in cyber security with the University of Sheffield. His research interests include the applications of artificial intelligence and machine learning in cyber security with special focus on targeting advanced persistent threat actors. He is an EU Marie-Curie Fellow.